

A Random Loop Model for Long Polymers

Manfred Bohn* and Dieter W. Heermann
*Institute of Theoretical Physics, University of Heidelberg,
Philosophenweg 19, D-69120 Heidelberg, Germany*

Roel van Driel
*Swammerdam Institute for Life Sciences, University of Amsterdam,
BioCentrum Amsterdam, Kruislaan 318, 1098 SM Amsterdam, The Netherlands*
(Dated: February 1, 2008)

Remarkably little is known about the higher-order folding motifs of the chromatin fibre inside the cell nucleus. Folding depends among others on local gene density and transcriptional activity and plays an important role in gene regulation. Strikingly, at fibre lengths above 5 to 10 Mb the measured mean square distance $\langle R^2 \rangle$ between any two points on the chromatin fibre is independent of polymer length. We propose a polymer model that can explain this levelling-off by means of random looping. We derive an analytical expression for the mean square displacement between two arbitrary beads. Here the average is taken over the thermal ensemble with a fixed but random loop configuration, while quenched averaging over the ensemble of different loop configurations – which turns out to be equivalent to averaging over an ensemble of random matrices – is performed numerically. A detailed investigation of this model shows that loops on all scales are necessary to fit experimental data.

PACS numbers: 87.15.-v, 87.15.Aa, 82.35.Pq, 82.35.Lr

Keywords: Looped Polymers, Biopolymers, Random Matrices, Random Walk

I. INTRODUCTION

The folding behaviour of DNA inside the cell nucleus has been subject to extensive studies. The diploid human genome consists of about 2 m of double stranded DNA, which is wrapped around histone protein octamers, forming nucleosomes every 180 to 250 bp. This chromatin fibre has a length of about 50 cm and is packaged inside a cell nucleus of about 10 μm diameter. Despite its tight folding, the DNA is readily accessible to large numbers of proteins required for transcription, replication and DNA repair. The first stages of compaction are well-known (see ref. [1] for a review): The DNA double strand is wrapped around histone octamers in a 1-and-3/4 left-handed superhelical turn forming the so-called nucleosomes, which are connected by stretches of linker DNA. This nucleosome-linker complex has a diameter of about 10 nm and is called 10 nm chromatin fibre. It has been shown *in vitro* that this fibre in turn condenses to a fibre of 30 nm diameter, whose structure is still under discussion [2, 3, 4].

Remarkably little is known about the higher-order folding of the chromatin fibre inside the human interphase nucleus [5]. It is not possible to follow the chromatin fibre in the interphase nucleus by imaging techniques. Therefore indirect approaches are being used to study chromatin folding. One method that has been applied by several groups is fluorescence in situ hybridization (*FISH*). By labelling chromatin in the intact cell with pairs of

FISH-markers one can measure the mean squared physical distance $\langle R^2 \rangle$ between these markers as a function of the genomic distance d . A recent study has shed light on the folding at larger length scales [6]. It was shown that there is a levelling-off in the physical vs. genomic distance plot at genomic distances longer than a few mega base pairs (Mb), so that approximately $\langle R^2 \rangle \sim O(1)$ for large d . At shorter length scales, i.e. up to a few Mb, the folding behaviour is best described within the globular state model, where the mean square physical distance scales like

$$\langle R^2 \rangle \sim d^{2/3} \quad . \quad (1)$$

The unexpected behaviour at larger genomic distances called for further explanation.

In the past, there have been several attempts to explain the folding motifs of chromatin in interphase cells using polymer models [7, 8, 9, 10]. One approach has been to model the chromatin fibre as a random walk in confined geometry [7]. Other data have been interpreted in terms of a Random-Walk/Giant-Loop (RWGL) model suggesting that there is a random walk backbone with regularly placed loops of about 3 Mb [8, 9] to explain different observed folding regimes on large and small scales. Another model, the Multi-Loop-Subcompartment (MLS) model [10, 11], proposes the existence of a rosette-like structure with 120 kb loops. None of these models presented so far is able to explain the levelling-off observed in the recent study.

It has been shown that the formation of loops of different sizes plays an important role in gene regulation and gene expression. One system that has been extensively studied is the β -globin locus [12, 13]. Here chromatin

*Electronic address: bohn@tphys.uni-heidelberg.de

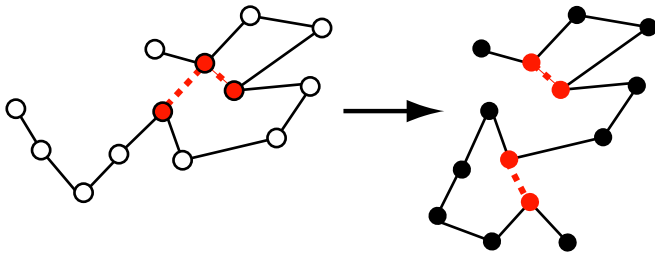


FIG. 1: The Random Loop Model averages over (a) the thermal disorder and (b) over the possible configurations of loops. Here one can see two possible configurations of loops.

expression is controlled by the formation of loops bringing different regulatory elements of the locus in physical contact. In this system the loop sizes are in the order of 10 kb. The formation of these loops is dynamic: different genes in the locus interact with the control locus during development in a mutually exclusive way, correlated with their expression. Loops that link promoter and enhancer complexes have been found of up to 3 Mb [14, 15]. Loops of up to several tens Mb have been associated with the formation of transcription factories, which bring together transcriptionally active genes [15, 16]. In all cases, chromatin loop formation is a dynamic process [17].

Based on these observations we propose a general polymer model which is able to explain the levelling-off behaviour observed in experiment where the mean squared physical distance scales like $\langle R^2 \rangle \sim O(1)$ at genomic distances above 1-2 Mb. The model takes into account the looping of the polymer, i.e. the chromatin fibre. The backbone of our polymer is formed by a random walk chain. For reasons of mathematical tractability we do not introduce excluded volume interactions. While Sachs *et al.* assumed loops of uniform size and fixed positions of the loop attachment points [9], we allow loops to have random polymer length and attachment points. We are interested in the average conformational properties of the system, mainly the mean square distance between two beads of the chain. We derive an analytical result for the average over the thermal disorder of one specific configuration of loops. The randomness of the loops gives rise to average over all different loop configurations. Numerical methods are used to get the quenched average over the different configurations of loops (see fig. 1 for two possible configurations). This average corresponds to the cell-to-cell variation of the measurements. We compare our model to the experimental data and investigate which loops are necessary to get a levelling-off where $\langle R^2 \rangle \sim O(1)$.

II. THEORY

A. Polymer models

In the past, experimental data from *FISH* measurements has been compared quite successfully to polymer models. Mainly three basic polymer models are used: (a) the random walk (RW) model where it is assumed that the orientations between chain segments are completely uncorrelated and no volume interactions are taken into account, (b) the self-avoiding walk model (SAW) takes excluded volume into account, while (c) the globular state (GS) model furthermore includes temperature-dependent attractive interactions (cf. ref. [18, 19]). One characteristic feature of a polymer model is the mean squared end-to-end distance $\langle R_N^2 \rangle$. For the basic models above, the end-to-end distance scales like

$$\langle R_N^2 \rangle = b^2 N^{2\nu} \quad (2)$$

in the limit of large N , where l is the linker length, N the chain length and ν a constant depending on the model used: $\nu = 0.5$ for the RW, $\nu \approx 0.588$ for the SAW and $\nu = 1/3$ for the GS. Assuming that eq. (2) also holds for intra-chain distances in the limit of large N , one can compare the experimental data to these models [6]. Relation (2) is no longer valid as soon as we introduce long-range interactions along the chain, such as chromatin looping.

Polymer modelling makes use of a coarse-graining approach, i.e. we divide the chromatin fibre into N equal subunits of length b . The subunits itself are assumed to be uncorrelated, i.e. they can freely rotate around each other. Such a description of a biological polymer is correct when we make N sufficient small so that b is larger than the persistence length of chromatin, i.e., the bending energy vanishes on this length scale.

B. General expression for the average over the thermal ensemble

To set up our Random Loop Model we first derive a general expression for the mean square distance between two arbitrary beads of the chain where harmonic interactions between all beads of the chain are allowed. The spatial positions of the chain's beads are denoted by $\mathbf{x}_0, \dots, \mathbf{x}_N$ and $\mathbf{x}_i = (x_i, y_i, z_i)$. For reasons of mathematical tractability we consider a phantom chain (random walk) where no excluded volume is taken into account. The Gaussian chain that forms the backbone of our model is characterized by the following potential,

$$U_{\text{Gaussian}} = \frac{\kappa}{2} \sum_{j=1}^N \| \mathbf{x}_j - \mathbf{x}_{j-1} \|^2 .$$

In addition to the random walk backbone, we allow each bead to interact with any other via harmonic potential,

resulting in an interaction term in the potential

$$U = U_{\text{Gaussian}} + \frac{1}{2} \sum_{\substack{i < j \\ |i-j| > 1}}^N \kappa_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

where $\kappa_{ij} = \kappa_{ji}$ are the spring constants for the loop attachment points. The spring constants κ_{ij} are given here and in the following in reduced units (comprising the term $1/k_B T$). Right now we keep them arbitrary but they will be randomly chosen later within the model. This potential has already been proposed by Sachs *et al.* [9] but not been elaborated further. It can be rewritten in the form

$$U = \frac{1}{2} \sum_{i < j}^N \kappa_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \frac{1}{4} \sum_{\substack{i, j=0 \\ j \neq i}}^N \kappa_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (3)$$

where $\kappa_{ij} = \kappa$ for $|i - j| = 1$.

The probability density for a bead conformation $(\mathbf{x}_0, \dots, \mathbf{x}_N)$ in the canonical ensemble is given by the Boltzmann factor

$$P(\mathbf{x}_0, \dots, \mathbf{x}_N) = C \exp\left(-\frac{U}{k_B T}\right) \quad (4)$$

where C is a normalization constant and $U = U(\mathbf{x}_0, \dots, \mathbf{x}_N)$ is the total potential energy of the chain.

We now eliminate the degrees of freedom stemming from the translational invariance of the problem by setting $\mathbf{x}_0 = \mathbf{0}$ (the absolute position of the chain in space is irrelevant for distances between beads).

Due to the Gaussian character of the probability density G , the spatial dimensions factorize,

$$P(\mathbf{x}_1, \dots, \mathbf{x}_N) = P_1(x_1, \dots, x_N) \cdot P_1(y_1, \dots, y_N) \cdot P_1(z_1, \dots, z_N)$$

and we can concentrate on the one-dimensional density function P_1 . By an easy calculation omitted here, we can rewrite the one-dimensional probability density P_1 as follows

$$P_1(x_1, \dots, x_N) = C_1 \exp\left(-\frac{1}{2} \mathbf{X}^T K \mathbf{X}\right) \quad (5)$$

where $\mathbf{X} = (x_1, \dots, x_N)^T$ and

$$K = \begin{pmatrix} \sum_{\substack{j=0 \\ j \neq 1}}^N \kappa_{1j} & -\kappa_{12} & \dots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j \neq 2}}^N \kappa_{2j} & \dots & -\kappa_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \dots & \sum_{\substack{j=0 \\ j \neq N}}^N \kappa_{Nj} \end{pmatrix} \quad (6)$$

Up to now we have not made any assumptions concerning the spring constants (i.e. basically the matrix entries)

κ_{ij} . In the following we only assume that K is a *symmetric* and *regular* matrix. So P_1 in eq. (5) turns out to be a multivariate normal distribution with mean $\mu = 0$ and covariance matrix $\Sigma = K^{-1}$. The marginal distribution for two arbitrary beads I and J ,

$$P(x_I, x_J) = \int \dots \int \prod_{\substack{i=1 \\ i \neq I, J}}^N P(x_1, \dots, x_N) \quad (7)$$

can be evaluated by standard methods for normal distributions. Going back to three dimensions we obtain after some basic integral evaluations the joint probability density for the distance between two beads I and J ,

$$P(\|\mathbf{x}_I - \mathbf{x}_J\|) \equiv P(r_{IJ}) = \tilde{C} r_{IJ}^2 \exp\left[-\frac{1}{2} \frac{1}{\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}} r_{IJ}^2\right].$$

Here

$$\Sigma = K^{-1} = (\sigma_{ij})_{i,j}$$

and \tilde{C} is the normalization constant. Using

$$\Gamma = \frac{1}{2} \frac{1}{\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}}$$

and calculating the correct normalization we obtain

$$P(r_{IJ}) = \frac{4}{\sqrt{\pi}} \Gamma^{\frac{3}{2}} r_{IJ}^2 \exp[-\Gamma r_{IJ}^2] \quad (8)$$

and finally

$$\begin{aligned} \langle r_{IJ}^2 \rangle_{\text{thermal}} &= \int r_{IJ}^2 P(r_{IJ}) dr_{IJ} \\ &= \frac{3}{2\Gamma} = 3(\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}). \end{aligned} \quad (9)$$

The bracket delimiters here denote the average over the thermal ensemble of $N + 1$ beads interacting via a given, but fixed harmonic potential.

C. The Random Loop Model

The last section was dedicated to the derivation of a general formula for the mean square displacement between two arbitrary beads of the chain where each bead may interact with any other via harmonic potential. This quantity turned out only to depend on the matrix K , or more accurately speaking, on its inverse. The matrix K contains all information about the interactions. Now we want to specify this matrix. Our model assumes the chromatin fibre to have a random walk backbone, meaning that $\kappa_{ij} = \kappa$ with $|i - j| = 1$. Furthermore the chromatin forms loops whose size and positions are randomly distributed along the chain. On a more general footing we

can restrict the possible loop sizes ℓ to a certain range $[l_1, l_2]$. Within this range all loops are chosen randomly by setting:

$$\kappa_{ij} = \begin{cases} \kappa & \text{with probability } \mathcal{P} \\ 0 & \text{with probability } 1 - \mathcal{P} \end{cases}, \quad \text{if } l_1 \leq |i - j| \leq l_2$$

$$\kappa_{ij} = 0 \quad \text{otherwise}$$

Note that we can set $\kappa = 1$ as it only scales the mean square displacement in eq. (9). Thus our model has two adjustable parameters, namely the chain length N and the probability \mathcal{P} .

The resulting matrices K represent an ensemble of diagonally dominated band random matrices and each matrix of this ensemble represents a loop configuration. This ensemble of random matrices has been investigated recently [20]. We are interested in the ensemble average of the mean square displacement, i.e. in the quantity

$$\begin{aligned} \langle r_{IJ}^2 \rangle &= \langle \langle r_{IJ}^2 \rangle_{\text{thermal}} \rangle_{\text{loops}} \\ &= 3 \left(\langle \sigma_{JJ} \rangle_{\text{loops}} + \langle \sigma_{II} \rangle_{\text{loops}} - 2 \langle \sigma_{IJ} \rangle_{\text{loops}} \right). \end{aligned}$$

This average is a quenched average and is equivalent to averaging over the ensemble of random matrices given by the above constraints. In sec. III B we also consider the case of the annealed ensemble and give an explanation why we use the quenched one here.

As our model already assumes that the chromatin fibre is translational invariant (as we do not take into account genomic content), we are only interested in the mean square displacement $\langle R_n^2 \rangle$ for two beads separated by $n = |i - j|$.

The average over the ensemble of random matrices cannot be performed analytically, so we have to use a representative subset of the ensemble and numerically calculate the inverse matrix and thereby the mean square displacement.

III. RESULTS

A. Comparison to experimental data

As noted earlier, our polymer model makes use of coarse-graining, since it is impossible to model such a long fibre in detail. Restrictions are given by computing time, which basically depends on the size of the matrix K . For our calculations we chose a matrix size of $N = 1000$ as a good compromise between computing time and not too coarse graining. Using a coarse-graining approach implies that we neglect details on a scale below the effective segment length being 150 kb in the following figures. Therefore we cannot resolve those loops that have been investigated in some gene-expression systems like the β -globin locus. As we are interested in large scale chromatin organization, it is justified to neglect these loops as they have no effect for the levelling-off at large

genomic distances, but only lead to a rescaling of the effective Kuhn length.

The first point of interest is which loops are necessary for the observed experimental behaviour. Do small loops (in the order of 100 kb to 1 Mb) already lead to the levelling-off, or are loops on all scales up to 80 Mb needed? Restricting the loop sizes to a range $\ell \in [1, s]$ only leads to a rescaling of the effective segment length, we still have $\langle R_n^2 \rangle \sim n$ (fig. 2a). In fig. 2b) we analyzed the ensemble where only large loops are allowed. Obviously, large loops are responsible for forcing the collapse of the chain, but the overall behaviour of the mean square physical distance does not fit the experimental data. Therefore loops on all scales are needed to obtain the levelling-off observed in experiment.

The characteristic features of the mean square displacement allowing loops on all scales can be seen in fig. 2c). At short contour lengths the mean square displacement grows similar to a random walk, but soon a levelling-off can be observed due to the attractive long-range interactions which is fairly $\sim O(1)$. While the contour length approaches N , the mean square displacement again rises to a random walk like behaviour. This is a chain end effect which is not of interest to us, as experiments only measure intra-chain distances. It is due to the construction of the loops, as the probability for having a loop with a larger size becomes increasingly small.

Thus, adding long-range interactions forcing the polymer to form loops yields completely different traits than a simple random walk or self-avoiding walk model. Note that the probabilities \mathcal{P} are chosen very small, meaning that a few loops suffice to obtain this levelling-off. The number of independent randomly chosen entries κ_{ij} is $\mathcal{C} = (N - 1)(N - 2)/2$ for a $N \times N$ -matrix and therefore the average number of loops per configuration is given by $\mathcal{C} \cdot \mathcal{P}$. With $\mathcal{P} = 4 \times 10^{-5}$ and $N = 1000$ one has an average of about 20 loops.

In fig. 3 the model is compared to the experimental data for different values of \mathcal{P} . Here one has to introduce two new scaling parameters, the segment length in physical units (e.g. nm) and the segment length in base pairs. The data is shown for a segment length of 300 nm and 150 kb. The latter is the size of the fluorescent markers used in experiments, therefore it does not make any sense to model on a more detailed scale. As mentioned above, using this coarse-graining approach all details on length scales smaller than 150 kb are neglected. The model can quite well explain the levelling-off at genomic distances above a few mega-base pairs as well as the rise at small genomic distances. As we have shown that on small genomic distances we have a globular-state-like behaviour [6], this random-walk-based model does not yield perfect results here.

In a recent publication [6] we already mentioned that plotting $\langle R_n^2 \rangle$ versus n is not a very sensitive method to check for the correctness of a model. Looking at the

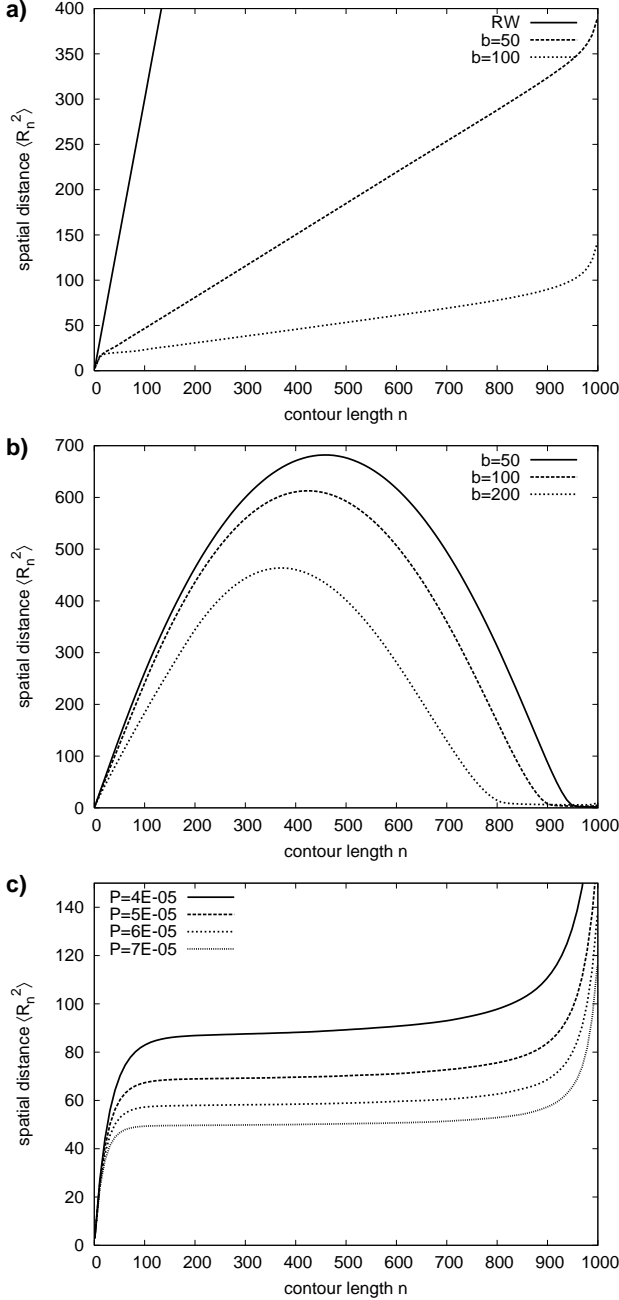


FIG. 2: Mean square displacement between two chain segments in relation to their contour length for the Random Loop Model for different allowed loop sizes. The chain length is always $N = 1000$. In figure **a**) only loops smaller than a certain size s are allowed. The basic scaling behaviour is still $\langle R_n^2 \rangle \sim n$ with a changed effective contour length compared to the free random walk. For this plot \mathcal{P} was chosen that the average number of loops per configuration is 100. Figure **b**) is for large loops where only loops of sizes ℓ in a range $[N - s, N]$ are allowed. While large loops seem to be responsible for the collapse of the chain, they alone cannot explain the experimental data. As in **a**) \mathcal{P} was chosen that the mean number of loops per configuration is 100. Figure **c**) shows the results for the situation where loops of all sizes are allowed. The levelling-off to $\langle R_n^2 \rangle \sim O(1)$ can already be achieved by a small number of loops

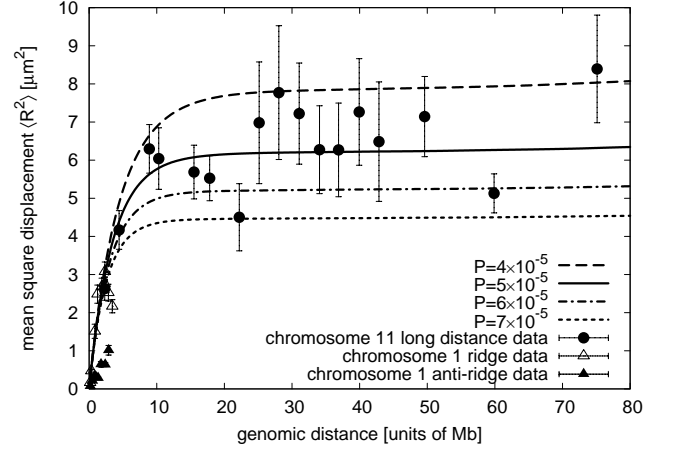


FIG. 3: Experimental data compared to the Random Loop Model. The data is taken from [6] and includes short and long distance sets. The results of the Random Loop Model are shown for $N = 1000$ and different values of \mathcal{P} .

cumulant relation between higher-order moments,

$$c_4 = \frac{\langle R^4 \rangle}{\langle R^2 \rangle^2} \quad (10)$$

gives much stronger evidence, as this expression is related to the distribution of the distances and not only its average value. Furthermore it has the advantage that the physical length scale divides out. Eq. (10) can be easily evaluated for a Gaussian Chain, where $c_4^{\text{RW}} = 5/3$. To obtain the value of the cumulant relation for a self-avoiding walk one has to use the expression for the distance probability density obtained by scaling arguments [21, 22],

$$P_{\text{SAW}}(R_N) = AR_N^{\mu+2} \exp\left(-D R_N^{\frac{1}{1-\nu}}\right), \quad \mu = 0.28 \quad (11)$$

Numerical integration gives $c_4^{\text{SAW}} \approx 1.506$, so both RW and SAW yield a constant expression. Fig. 4 shows that this constant has a value significantly below the fluctuations of the data. Here our model is in better agreement with experiments. We should point out here the importance of averaging over the disorder of loops. The cumulant expression c_4 only averaged over the thermal ensemble given by equations (8) and (9) is the same as for a pure random walk, namely $5/3$. It is the average over the loop configurations that changes this behaviour, bringing it in better agreement with the data.

B. Annealed vs. quenched average

In systems with disorder one has to perform averages both over a set of statistical variables and over a set of random variables representing the disorder [23]. In the case of the Random Loop Model the partition sum \mathcal{Z}

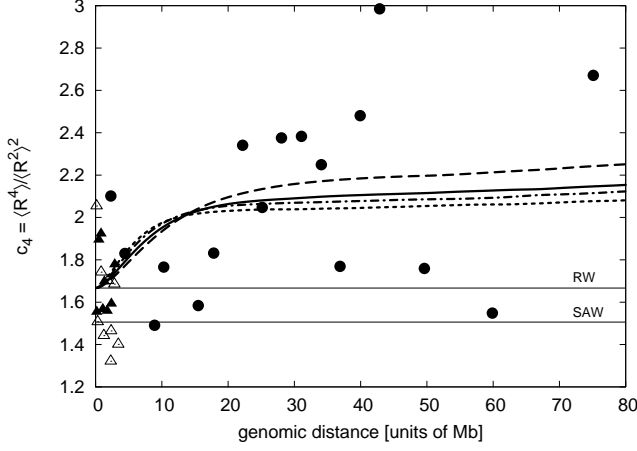


FIG. 4: The cumulant expression $c_4 = \langle R^4 \rangle / \langle R^2 \rangle^2$ of the experimental data compared to SAW, RW (straight lines) and Random Loop Model (dashed lines). For the legend, see fig. 3. The data shows significant differences to a RW and a SAW (horizontal lines), and is fluctuating around the cumulant of the Random Loop Model.

depends on the statistical variables $\mathbf{x}_1, \dots, \mathbf{x}_N$ and a set of random variables representing the disorder $\{\kappa_{ij}\}$ with $|i - j| > 1$. Above we have performed the average first over the statistical variables, while the average over the disorder has been taken over the quantity $\langle r_{ij} \rangle^2$. This corresponds to a process of quenched averaging. Using this method we assume that the time the cell needs to go into a new disorder configuration is much longer than the time needed for the cell to come into thermal equilibrium, $\tau_{\text{eq}} \ll \tau_{\text{dis}}$. At least to our knowledge, such time scales are not known inside the cell nucleus. Time evolution measurements cannot be performed on a cell using *FISH* markers as the cell has to be fixated before applying imaging techniques. Different configurations can only be observed by looking at different cells. It seems to us more reasonable to use a quenched type of average for comparison to biological data, as the loops are functional complexes which have to persist a while in order to properly fulfill their tasks.

Nevertheless, it is interesting to consider the other case, the annealed ensemble. This type of average should be applied if $\tau_{\text{dis}} \ll \tau_{\text{eq}}$. The annealed average of the partition sum can be written as

$$\langle \mathcal{Z} \rangle_{\text{ann}} = \sum_{\{\kappa_{ij}\}} \mathcal{Z}(\{\mathbf{x}_k\}, \{\kappa_{ij}\}) p(\{\kappa_{ij}\}) \quad , \quad (12)$$

where the sum is over all possible configurations of disorder, and p is the probability of one such configuration. We restrict our calculation to the case that loops of all sizes are allowed and that the spring constant is κ for adjacent beads as well as for loops. Assuming that the κ_{ij} are i.i.d. Bernoulli as before, the average over the

disorder can be carried out exactly,

$$\langle \mathcal{Z} \rangle_{\text{ann}} = \int d\mathbf{x}_1 \dots d\mathbf{x}_N \exp(-U_{\text{Gaussian}}) \times \prod_{i < j-1} \left[\mathcal{P} \left(e^{-\frac{1}{2}\kappa \|\mathbf{x}_i - \mathbf{x}_j\|^2} - 1 \right) + 1 \right] \quad .$$

Introducing the effective potential

$$U_{\text{eff}} = \frac{1}{2}\kappa \sum_{i=0}^{N-1} r_{i,i+1}^2 - \sum_{|i-j|>1} \log \left[1 + \mathcal{P} \left(e^{-\frac{1}{2}\kappa r_{ij}^2} - 1 \right) \right] \quad ,$$

where $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$, we can rewrite the partition sum as

$$\langle \mathcal{Z} \rangle_{\text{ann}} = \int \int d\mathbf{x}_1 \dots d\mathbf{x}_N \exp(-U_{\text{eff}}) \quad (13)$$

The effective potential has two parts: Adjacent beads with $|i - j| = 1$ keep their attractive harmonic potential, while *all* non-adjacent beads interact via a pairwise attractive potential $V(r)$. This potential is characterized by a minimum at $r = 0$, while for large r it reaches a plateau at $V(r \rightarrow \infty) = -\log(1 - \mathcal{P})$. In a low temperature approximation, a series expansion around $r = 0$ up to second order gives

$$V(r) = \frac{1}{2}\mathcal{P}\kappa r^2 \quad (14)$$

– a harmonic potential with effective spring constant $\mathcal{P}\kappa$.

The partition sum in (13) cannot be evaluated analytically and therefore we do not obtain an expression for the mean square displacement in the annealed case. One could obtain results using extensive and time-consuming MD or MC simulations. It will be left for future investigations.

C. Limiting cases without disorder

In most cases one cannot solve the model presented above analytically. Using the quenched ensemble one cannot calculate the average over the disorder, while using the annealed ensemble one cannot obtain the partition sum after having performed the disorder average. Therefore we calculated sample averages for the quenched case above. There are two special cases where the model can be solved exactly. These are the limiting cases where no disorder is present. $\mathcal{P} = 0$ is the situation of a normal Gaussian Chain with spring constant κ . It is well known that the mean square distance between two beads separated by n monomers is given by $\langle R_n^2 \rangle = \frac{3}{\kappa}n$. The other limit, $\mathcal{P} = 1$, corresponds to a fully connected network of beads. Assuming that all beads interact with spring constant κ , we can solve this problem analytically. Here we basically do not deal with a linear chain any more. The interaction matrix $K = (k_{ij})_{i,j}$ in this case writes

$$k_{ij} = \begin{cases} N\kappa & \text{for } i = j \\ -\kappa & \text{for } i \neq j \end{cases} \quad (15)$$

By an easy calculation one can show that the inverse matrix is given by

$$\sigma_{ij} = \begin{cases} \frac{2}{(N+1)\kappa} & \text{for } i = j \\ \frac{1}{(N+1)\kappa} & \text{for } i \neq j \end{cases} \quad (16)$$

Recall our definition of the chain at the beginning of sec. II B: Although we have an $N \times N$ -matrix our chain has $N + 1$ beads, as we set $\mathbf{x}_0 \equiv 0$. Inserting into eq. (9) yields

$$\langle R_n^2 \rangle \equiv \langle r_{ij}^2 \rangle = \frac{3}{(N+1)\kappa/2} \quad (17)$$

Within this system two beads are interacting with an effective harmonic potential with $\kappa_{\text{eff}} = (N+1)\kappa/2$.

Of major interest is the case where $\mathcal{P} = 1$, but where adjacent beads interact with a different spring constant than loops, i.e. $\kappa_{ij} = \kappa$ for $|i - j| = 1$ and $\kappa_{ij} = \hat{\kappa}$ for $|i - j| > 1$. We were not able to solve this case analytically. One might take this system as a model for the low-temperature limit of the annealed case in eq. (14) where κ is replaced by an effective interaction $\hat{\kappa} = \mathcal{P}\kappa$. On a more general footing this case might also be regarded as a model for a system where the random attraction with probability \mathcal{P} and loop spring constant κ has been replaced by an average attraction with probability $\mathcal{P} = 1$ and loop spring constant $\mathcal{P}\kappa$. It is clear a priori that such a potential will lead to a collapse of the chain, as all beads are interconnected. In fig. 5 we chose $\hat{\kappa} = \kappa = 1$ and $\mathcal{P} = 4 \times 10^{-5}$ as the reference curve. In comparison with the case of average attraction ($\mathcal{P} = 1, \kappa = 1, \hat{\kappa} = \mathcal{P}$) the levelling-off is much less pronounced. Of course it is possible to come into close agreement with the reference curve by choosing another interaction constant. For our reference curve one would have to lower $\hat{\kappa}$ by about one order of magnitude, corresponding to $\mathcal{P} \sim 2 \times 10^{-6}$ (< 1 loop per chain!). Although one could fit the data with these averaged attraction potential, we see no biological reason for such a potential to exist in the cell.

IV. CONCLUSIONS

In this paper we present a polymer model for the folding of the chromatin fibre in the interphase nucleus, based on recent experimental results [6]. These show that the chromatin fibre inside the human cell nucleus is ruled by two different folding regimes: At small genomic distances, i.e. below ~ 3 Mb the data can be explained well with a globular state polymer. At larger genomic distances there is a levelling-off to a scaling behaviour $\langle R^2 \rangle \sim O(1)$. This indicates the presence of long-range interactions. In agreement with the experimental findings we propose the existence of loops to be responsible for this levelling-off. The main aspect of our model is the assumption of loops of various sizes and positions, where we average not only over the thermal ensemble but also

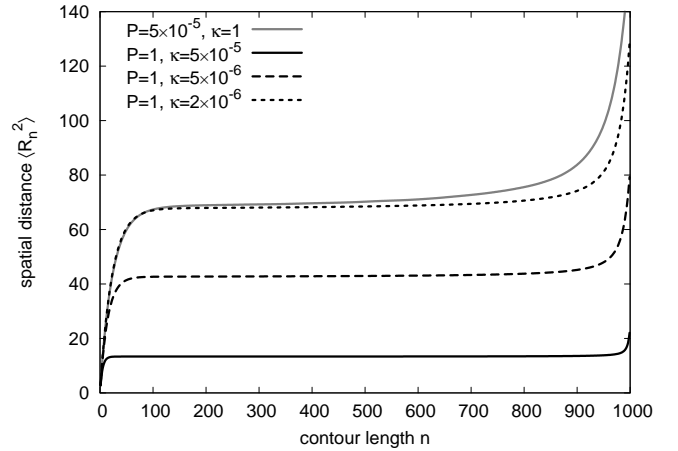


FIG. 5: The Random Loop Model (RLM) compared to a system where the random attraction (setting $\kappa_{ij} = 1$ with probability \mathcal{P}) has been replaced by an average attraction (setting all $\kappa_{ij} = \mathcal{P}\kappa$ for $|i - j| > 1$). Shown are the RLM reference curve for $\mathcal{P} = 5 \times 10^{-5}$ (grey line), the corresponding system with average attraction (black line) and two systems with smaller average attraction.

over different configurations of loops. Keeping the looping probability \mathcal{P} constant in the ensemble. Eq. (9) gives a general expression for the mean square physical distance within the thermal ensemble between two arbitrary chosen beads, when each bead is allowed to interact with any other via harmonic potential. We show that $\langle R^2 \rangle$ only depends on the interaction matrix K , which we have introduced in (6). Within the Random Loop Model the matrix entries (equivalent to random loops) are chosen in a random manner to be $-\kappa$ with probability \mathcal{P} or 0 otherwise. The average over the ensemble of loop configurations then turns out to be equivalent to averaging over the specific ensemble of random matrices with fixed \mathcal{P} . We show that random loops on all length scales explain the levelling-off observed in experiment, while restricting the loop sizes to only large loops (in the order of 50 - 100 Mb) or small loops (100 kb - 10 Mb) does not fit the data.

In contrast to the Random-Walk/Giant-Loop model [9] we do not assume fixed-size and regular placed loops. The average over the ensemble of different loop configurations turns out to be necessary to obtain the levelling-off observed in experiment. Its importance also becomes apparent in fig. 4, where the RWGL model would yield the same result as the random walk, namely a constant: $c_4^{\text{RWGL}} = 5/3$.

In a recent paper [6] we have shown that on short genomic distances the globular state fits the experimental data best compared to a self-avoiding and random walk. In the present study we have neglected the effect of excluded volume. Therefore on short genomic distances our model will explain the data inadequately as it shows random-walk-like behaviour by design on short contour lengths, because for a globular state polymer the expo-

nent in eq. (2) is $\nu = 1/3$, while for a random walk polymer it is $\nu = 1/2$. On larger genomic distances the existence of a small number of loops can explain the levelling-off without the need for introducing excluded volume interactions. The role of excluded volume inside the cell nucleus on the spatial chromatin properties remains an open question, because of the impact of an enzyme called topoisomerase II, which is capable of cutting the DNA double strand in order to let another strand pass through it. This event might well give rise to a random-walk-like statistics, dependent on the frequency of these events.

The model so far is able to explain some basic features

of DNA folding revealed in recent experiments [6]. Future work on this model might encompass for example the integration of excluded volume interactions and its effect on the mean square displacement.

Acknowledgments

M. Bohn gratefully acknowledges funding from the Landesgraduiertenförderung Baden-Württemberg.

-
- [1] H. Schiessel, W. M. Gelbart, and R. Bruinsma, *Biophys. J.* **80**, 1940 (2001).
 - [2] J. Widom, *Annu Rev Biophys Chem* **18**, 365 (1989).
 - [3] K. van Holde and J. Zlatanova, *J. Biol. Chem.* **270**, 8373 (1995).
 - [4] K. van Holde and J. Zlatanova, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10548 (1996).
 - [5] R. Horowitz-Scherer and C. Woodcock, *Chromosoma* **115**, 1 (2006).
 - [6] J. Mateos-Langerak, O. Giromus, W. de Leeuw, M. Bohn, G. Kreth, D. W. Heermann, R. van Driel, and S. Goetze, [arXiv:0705.1656 \[q-bio.GN\]](https://arxiv.org/abs/0705.1656) (2007).
 - [7] P. Hahnfeldt, J. E. Hearst, D. J. Brenner, R. K. Sachs, and L. R. Hlatky, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7854 (1993).
 - [8] H. Yokota, G. van den Engh, J. Hearst, R. Sachs, and B. Trask, *J. Cell Biol.* **130**, 1239 (1995).
 - [9] R. Sachs, G. Engh, B. Trask, H. Yokota, and J. Hearst, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2710 (1995).
 - [10] C. Münkler and J. Langowski, *Phys. Rev. E* **57**, 5888 (1998).
 - [11] C. Münkler, R. Eils, S. Dietzel, D. Zink, C. Mehring, G. Wedemann, T. Cremer, and J. Langowski, *J. Mol. Biol.* **285**, 1053 (1999).
 - [12] R.-J. Palstra, B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat, *Nat. Genet.* **35**, 190 (2003).
 - [13] W. de Laat and F. Grosveld, *Chromosome Res.* **11**, 447 (2003).
 - [14] M. Petrascheck, D. Escher, T. Mahmoudi, C. P. Verrijzer, W. Schaffner, and A. Barberis, *Nucleic Acids Res.* **33**, 3743 (2005).
 - [15] P. Fraser, *Curr. Opin. Genet. Dev.* **16**, 490 (2006).
 - [16] P. Fraser and W. Bickmore, *Nature* **447**, 413 (2007).
 - [17] P. R. Cook, *Nat. Genet.* **32**, 347 (2002).
 - [18] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, 1994).
 - [19] P.-G. de Gennes, *Scaling concepts in polymer physics* (Ithaca, N.Y., Cornell University Press, 1979).
 - [20] D. W. Heermann and M. Bohn, [arXiv:0705.1241v1 \[cond-mat.stat-mech\]](https://arxiv.org/abs/0705.1241v1) (2007).
 - [21] M. Baiesi, E. Carlon, Y. Kafri, D. Mukamel, E. Orlandini, and A. L. Stella, *Phys. Rev. E* **67**, 021911 (2003).
 - [22] M. Fisher, *J. Chem. Phys.* **44**, 616 (1966).
 - [23] K. Binder and A. P. Young, *Rev. Mod. Phys.* **58**, 801 (1986).